B10

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/08

A2

(11) International Publication Number: WO 98/30717

(43) International Publication Date: 16 July 1998 (16.07.98)

(21) International Application Number: PCT/EP97/07134

(22) International Filing Date: 2 December 1997 (02.12.97)

(30) Priority Data:
60/032,069 2 December 1996 (02.12.96) US

(71) Applicant *(for all designated States except US)*: BIOCEM S.A. [FR/FR]; Campus Universitaire des Cézeaux, 24, avenue des Landais, F-63170 Aubière (FR).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: MURIGNEUX, Alain [FR/FR]; Biocem S.A., Campus Universitaire des Cézeaux, 24, avenue des Landais, F-63170 Aubière (FR).

(74) Agent: BREESE-MAJEROWICZ; 3, avenue de l'Opéra, F-75001 Paris (FR).

(81) Designated States: AU, CA, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: VEGETAL SEQUENCES INCLUDING A POLYMORPHIC SITE AND THEIR USES

(57) Abstract

A nucleic acid segment comprising at least 10 contiguous nucleotides from a vegetal sequence including a polymorphic site; or the complement of the segment.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

VEGETAL SEQUENCES INCLUDING A POLYMORPHIC SITE AND THEIR USES

The genomes of all organisms undergo spontaneous mutation in the course of their continuing evolution generating variant forms of progenitor sequences (Gusella, *Ann, Rev. Biochem.* 55, 831-854 (1986)). The variant form may confer an evolutionary advantage or disadvantage relative to a progenitor form or may be neutral. In some instances, a variant form confers a lethal disadvantage and is not transmitted to subsequent generations of the organism. In other instances, a variant form confers an evolutionary advantage to the species and is eventually incorporated into the DNA of many or most members of the species and effectively becomes the progenitor form. In many instances, both progenitor and variant form(s) survive and co-exist in a species population. The coexistence of multiple forms of a sequence gives rise to polymorphisms.

Several different types of polymorphism have been reported. A restriction fragment length polymorphism (RFLP) means a variation in DNA sequence that alters the length of a restriction fragment as described in Botstein et al., Am. *J. Hum. Genet.* 32, 314-331 (1980). The restriction fragment length polymorphism may create or delete a restriction site, thus changing the length of the restriction fragment. RFLPs have been widely used in human and animal genetic analyses (see WO 90/13668; WO 90/11369; Donis-Keller, *Cell* 51, 319-337 (1987); Lander et al., *Genetics* 121, 85-99 (1989)). When a heritable trait can be linked to a particular RFLP, the presence of the RFLP in an individual can be used to predict the likelihood that the animal will also exhibit the trait.

Other polymorphisms take the form of short tandem repeats (STRs) that include tandem di-, tri- and tetra-nucleotide repeated motifs These tandem repeats are also referred to as variable number tandem repeat (VNTR) polymorphisms. VNTRs have been used in identity and

paternity analysis (US 5,075,217; Armour et al , *FEBS Lett.* 307, 113-115 (1992); Horn et al., WO 91/14003; Jeffreys, EP 370,719), and in a large number of genetic mapping studies.

Other polymorphisms take the form of single nucleotide variations between individuals of the same species. Such polymorphisms are far more frequent than RFLPs, STRs and VNTRs. Some single nucleotide polymorphisms occur in proteincoding sequences, in which case, one of the polymorphic forms may give rise to the expression of a defective or other variant protein. Other single nucleotide polymorphisms occur in noncoding regions. Some of these polymorphisms may also result in defective or variant protein expression (e.g., as a result of defective splicing). Other single nucleotide polymorphisms have no phenotypic effects. Single nucleotide polymorphisms can be used in the same manner as RFLPs, and VNTRs but offer several advantages. Single nucleotide polymorphisms occur with greater frequency and are spaced more uniformly throughout the genome than other forms of polymorphism. The greater frequency and uniformity of single nucleotide polymorphisms means that there is a greater probability that such a polymorphism will be found in close proximity to a genetic locus of interest than would be the case for other polymorphisms. Also, the different forms of characterised single nucleotide polymorphisms are often easier to distinguish that other types of polymorphism (e.g., by use of assays employing allele-specific hybridization probes or primers).

Despite the increased amount of nucleotide sequence data being generated in recent years, only a minute proportion of the total repository of polymorphisms has so far been identified. The paucity of polymorphisms hitherto identified is due to the large amount of work required for their detection by conventional methods. For example, a conventional approach to identifying polymorphisms might be to sequence the same stretch of oligonucleotides in a

population of individuals by didoxy sequencing. In this type of approach, the amount of work increases in proportion to both the length of sequence and the number of individuals in a population and becomes impractical for large stretches of DNA or large numbers of subjects.

SUMMARY OF THE INVENTION

The invention provides nucleic acid segments containing at least 10, 15 or 20 contiguous bases from a vegetal fragment including a polymorphic site notably a single nucleotide polymorphism (SNP). In a particular embodiment, a vegetal fragment does not belong to the Cruciferae family.

The segments can be DNA or RNA, and can be double- or single-stranded. Some segments are 10-20 or 10-50 bases long. Preferred segments include a diallelic polymorphic site. In a preferred embodiment, the invention concerns nucleic acid segments from a fragment shown in Table I (corn).

The Invention further provides allele-specific oligonucleotides that hybridizes to a segment of a vegetal fragment, for example fragment in Table I. These oligonucleotides can be probes or primers. Also provided are isolated nucleic acid" comprising a sequence of Table I or the complement thereto, in which the polymorphic site within the sequence is occupied by a base other than the reference base shown in Table I.

The invention further provides a method of analyzing a nucleic acid from a subject. The method determines which base or bases is/are present at any one of the polymorphic vegetal sites for example of those of Table I. Optionally, a set of bases occupying a set of the polymorphic sites shown in Table I is determined. This type of analysis can be performed on a plurality of subjects who are tested for the presence of a phenotype. The presence or absence of phenotype can then be correlated with a base or

set of bases present at the polymorphic sites in the subjects tested.

DEFINITIONS

A nucleic acid, such an oligonucleotide, oligonucleotide can be DNA or RNA, and single- or double-stranded. Oligonucleotides can be naturally occurring or synthetic, but are typically prepared by synthetic means. Preferred nucleic acids of the invention include segments of DNA, or their complements including any one of the polymorphic sites shown in Table I. The segments are usually between 5 and 100 bases, and often between 5-10, 5-20, 10-20, 10-50, 20-50 or 20-100 bases. The polymorphic site can occur within any position of the segment. The segments can be from any of the allelic forms of DNA shown in Table I. Methods of synthesizing oligonucleotides are found in, for example, *Oligonucleotide Synthesis: A Practical ApproacA* (Gait, ed., IRL Press, Oxford, 1984).

Hybridization probes are oligonucleotides capable of binding in a base-specific manner to a complementary strand of nucleic acid. Such probes include peptide nucleic acids, as described in Nielsen et al., *Science* 254, 1497-1500 (1991).

The term primer refers to a single-stranded oligonucleotide capable of acting as a point of initiation of template-directed DNA synthesis under appropriate conditions (i.e., in the presence of four different nucleoside triphosphates and an agent for polymerization, such as, DNA or RNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. The appropriate length of a primer depends on the intended use of the primer but typically ranges from 15 to 30 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template A primer need not reflect the exact sequence of the template but must be sufficiently complementary to hybridize with a template. The term primer site refers to the area of the target DNA to which a primer

hybridizes. The term primer pair means a set of primers including a 5' upstream primer that hybridizes with the 5' end of the DNA sequence to be amplified and a 3', downstream primer that hybridizes with the complement of the 3' end of the sequence to be amplified.

Linkage describes the tendency of genes, alleles, loci or genetic markers to be inherited together as a result of their location on the same chromosome, and can be measured by percent recombination between the two genes, alleles, loci or genetic markers.

Polymorphism refers to the occurrence of two or more genetically determined alternative sequences or alleles in a population. A polymorphic marker or site is the locus at which divergence occurs. Preferred markers have at least two alleles, each occurring at frequency of greater than 1%, and more preferably greater than 10% or 20% of a selected population. A polymorphic locus may be a" small as one base pair. Polymorphic markers include restriction fragment length polymorphisms, variable number of tandem repeats (VNTR's), hypervariable regions, minisatellites, dinucleotide repeats, trinucleotide repeats, tetranucleotide repeats, simple sequence repeats, and insertion elements such as Alu. The first identified allelic form is arbitrarily designated as a the reference form and other allelic forms are designated as alternative or variant alleles. The allelic form occurring most frequently in a selected population is sometimes referred to as the wildtype form. Diploid organisms may be homozygous or heterozygous for allelic forms. A diallelic polymorphism has two forms. A triallelic polymorphism has three forms.

A single nucleotide polymorphism occurs at a polymorphic site occupied by a single nucleotide, which is the site of variation between allelic sequences. The site is usually preceded by and followed by highly conserved sequences of the allele (e.g., sequences that vary in 1QSS than 1/100 or 1/1000 members of the populations).

A single nucleotide polymorphism usually arises due to substitution of one nucleotide for another at the polymorphic site. A transition is the replacement of one purine by another purine or one pyrimidine by another pyrimidine. A transversion is the replacement of a purine by a pyrimidine or vice versa. Single nucleotide polymorphisms can also arise from a deletion of a nucleotide or an insertion of a nucleotide relative to a reference allele.

Hybridizations are usually performed under stringent conditions, for example, at a salt concentration of no more than 1 M and a temperature of at least 25°C For example, conditions of 5X SSPE (750 mM NaCl, 50 mM NaPhosphate, 5 mM EDTA, pH 7.4) and a temperature of 25-30°C are suitable for allele-specific probe hybridizations.

Nucleic acids of the invention are often in isolated form. An isolated nucleic acid means an object species that is the predominant species present (i.B., on a molar basis it is more abundant than any other individual species in the composition). Preferably, an isolated nucleic acid comprises at least about 50, 80 or 90 percent (on a molar basis) of all macromolecular species present. Most preferably, the object species is purified to essential homogeneity (contaminant specie" cannot be detected in the composition by conventional detection methods).

DESCRIPTION OF THE PRESENT INVENTION

I. Novel Polymorphisms of the Invention

The present application provides for example oligonucleotides containing polymorphic sequences isolated from graminae species for example maize. The invention also includes various methods for using those novel oligonucleotides to identify, distinguish, and determine the relatedness of individual strains or pools of nucleic acids from plants.

EXAMPLES

Example 1. Maize DNA extraction

DNA was extracted from maize lines as described in Rogers and Bendich (1988 Plant Mol Biol Manual A6 : 1-10) with modification described in Murigneux et al (1993 theo Appl Genet 86 : 837-842).

PCR amplification was done on six maize lines representing a wide range of genetic variability and including both european flint material and US dent germplasm. Those six maize lines have been choosen to maximize the genetic variability of cultivated maize. Doing so, optimize the chance of finding polymorphism in the allelic sequences. For example G1, an european flint line and G3, an US Corn Belt Stiff Stalk line, are completly unrelated. Their genetic distance (coefficient of dissimilarity) calculated with our standard approach (89 RFLP probe/enzyme combinations and Nei-li distance) is 0.69. This value is close to the maximum distance between two cultivated maize lines.

Among the 15 genetic distance between couple of these 6 lines : 8 are superior to 0.6, 6 superior to 0.5 and only one inferior to 0.5. This shows that the choice of the lines avoided as much as it was possible the potential redudancy (or similarity) of allele at the locus sequenced. With the same effort of sequencing we should therefore have collected the maximum number of polyphomism.

Genotypes :

G1=flint line
G2=flint line
G3=Dent line
G4=Dent line
G5=Dent line
G6=Dent line

Example 2. Choice of the markers

The markers have been chosen with the following criteria.

1. Selection of markers that give a single band in southern hybridization. This is to avoid as much as

possible the problems of duplicated sequences (very frequent in plants). If the same (or nearly the same) sequence occurs at several position in the genome (locus 1 and 2) and if the primers used to type the SNP found on locus 1 allow amplification of the sequence at the locus 2, the results of hybridization on the chips will be the addition of two markers pattern and therefore impossible to use.

2. Distribution on the genome : most of the genetic analysis in plant aim to characterize the whole genome (genetic variability evaluation, mapping quantitative trait-locus, back-cross assisted selection). The second criteria was therefore to choose markers nicely distributed over the 10 chromosomes (see Table A hereunder for map position).

3. Selection of gene coding for enzymes involved in the Carbone metabolism. Wx1, Ae1, Sh2, Bre1, Bt1, Ssu, Bt2 are involved in sugar-starch metabolism. Such a choice will allow to have a very fast characterization of the allelic variability (possibly linked to efficiency) of gene involved in this metabolism.

The following markers have been used : see Table A.

LEGEND OF TABLE A

Probe = name of the marker

COD = in-house code.

MAP Pos = map position, given by the bin location of the University of Missouri map (Maize Genetic Newsletter n°69 1995). Examples of reading the "MAP Pos" and "Prim" columns : 1.01-1.02 means that it is the core probe that delimit the bins 1 and 2 on chromosome 1

5.01 means that it is located in the bin 5.01 (on chromosome 5)

4 means that it is located on chromosome 4

S01F is the forward primer for probe 1

S01R is the reverse primer for probe 1

Genbank/ EMBL = Genbank/ EMBL number

TABLE A

Csnp1d (33 markers)

PROBE	COD	Map Pos	PRIM	SEQUENCES OLIGOS	Genbank/EMBL
UMC157	S01	1.01-1.02	S01F	CGCACGCACATTAGCTTTCG	G10822
			S01R	TGCAACCGAACAGGATCTGC	G10823
UMC76	S02	1.02-1.03	S02F	ATTATTCGGCGTCCAGCCCC	G10865
			S02R	TTACCAGCGGTGAGAGCTGC	G10866
UMC67	S03	1.05-1.06	S03F	CGTTCGTGTGGCATCAATCG	G10864
			S03R	CGACATCATCATCGGCAACC	G13173
UMC161	S06	1.10-1.11	S06F	CAGACCTTGGTTGGAGGCAAC	G10824
			S06R	TCGCTCCCCTTCTTCCTTCC	G10825
UMC53	S08	2.01-20.2	S08F	CGGACGTGATGCAAGTTTCG	G10851
			S08R	AGCGGCTCAAGCTCTCCATC	G10852
UMC131	S10	2.04-2.05	S10F2	TCCTTGGCACTCACGCTACC	G10816
			S10R2	AGCATGGGGGGCAACAACTC	G10817
UMC49	S12	2.08-2.09	S12F	CAGAGAGCCGTCTCGAATCG	G10845
			S12R	TTGATACTGCCGTCTGCCG	G10846
UMC102	S14	3.04-3.05	S14F	TGCTGTGCTGTCACATGGCG	G10801
			S14R	CTGGGTCGTCGTGCTTTGAG	G10802
UMC63	S16	3.08-3.09	S16F2	ACGCCCTGACAGAACCATCG	G10857
			S16R	TTGCTCACTCGTGGTCGTGG	G10857
Adh2	S17	4.03	S17F2	TGCCTGCTGCATCTCTAGCC	X02915
			S17R2	CAAGCCCGAAAATCGCCAC	X02915
UMC66	S19	4.06-4.07	S19F	TGGAGTGTCCAAAGACCGACC	G10862
			S19R	ACCAAAACGGGTGGTCTGCC	G10863
UMC90	S22	5.01	S22F	GCAGGTGAACAATGCTGCCC	G10870
			S22R	CCAAAAGGCGGAGAACCGAC	G10871
Ae1	S23	5.05	S23F	TCGCTGGGGTTTTAGCATTG	L08065
			S23R	CACTCGAACTCTGTTCAAGGCTTG	L08065
UMC59	S26	6.01-6.02	S26F	TCCAAAGCGAAAGCCTGATG	G10853
			S26R	TACGATGGCCGTGACCCTTC	G10854
UMC65	S27	6.03-6.04	S27F	TTCCAGCTTTCCTCGGCACC	G10860
			S27R	AGCAGCAAGAGCAGAGCGTG	G10861
UMC21	S28	6.04-6.05	S28F	TGCAGATGTGCCTTTCCTGTG	G10830
			S28R	CAGTGGATTCGCTCCCTTCTC	G10831
UMC132	S29	6.06-6.07	S29F	CGCACAGAGGCAGATGCAGC	G10824
			S29R	CGCTAGGCAGAGGTTCGAGC	G10819
UMC254	S33	7.03-7.04	S33F	CCGGGCGCAAAAGAATGTG	G10832
			S33R	AAGAAACCAGCACCAGCGGG	G10833
UMC80	S34	7.04	S34F	TCGCCTTTATCGGTGCAATG	G10867
			S34R	TGGAGCAAGCATGGAGATCG	G10868
BNL9-11	S38	8.01-8.02	S38F2	CGAGGGAATGTCATCAACCC	G10778
			S38R2	ACCAAAGCTCCTCAGCCAAG	G10779
UMC109	S42	9.00-9.01	S42F	GCACCGTCGTTTACCTCAAGC	G13177
			S42R	TAGCCATCATCAGCGGCGTG	G10807
Wx1	S43	9.02-9.03	S43F	CGTGCTACCTCAAGAGCAAC	X03935
			S43R	ACTTCACGGCGATGTACTTG	X03935
UMC95	S44	9.04-9.05	S44F	CACTCGGAAGTCGGAATCGC	G10872
			S44R	ACCTTCGCAGTGTTGCGGAC	G10872
CSU61	S45	9.05-9.06	S45F	TCTCCACGAATCCCACCGTC	T12691
			S45R	AAGGGAGGGAATCCTCTACCG	T12691
UMC130	S48	10.02-10.03	S48F	AAGGGGGAAGAAGGTCATC	G10814
			S48R	CGATGGCAACAACTACCAGTAG	G10815
CSU109	S53	2.09	S53F	GCTTTCGGTTCCGGATAGCG	T12721
			S53R	ACTGGGCCATCTCCGACCAG	T12721
UAZ77	S56	5.04	S56F2	GCAACCAACTGCAACATCGC	T18762
			S56R2	GAAGGAGCTCAAGGCCAAGG	T18762

Sh1	S57	9.01	S57F	TGCTGTTATTGCGTGCCGTG	X02382
			S57R2	AAGGTGGCACCAAGGCGTTC	X02382
Sh2	S63	3.09	S63F	TTCTTCACTGCACCCCGATG	M81603
			S63R	CTGCTCACTCTGCAATGCCC	M81603
Bre1	S65	6	S65F	AGCAGCAGATCAGGCACACC	U17897
			S65R	TTGAAGTTCGTTTCGGGCAC	U17897
Bt1	S66	5	S66F	GGCAAGGATCGGAGTTGCTC	M79333
			S66R	TAGCGTGGAGGACGTTCTGG	M79333
Ssu	S67		S67F	GCAAGCAAGCAAGCAGCGAG	D00170
			S67R	GACCCGAAGCAAAACCGAAC	D00170
Bt2	S71	4	S71F	TGCCGAAAAAGGTGGCATTC	Seq (Bae et al 1990)
			S71R	GCCCCCAATGTCGATTCAAC	

Example 3. PCR amplification

PCR amplification was done with primer designed using the DNA sequences of the markers listed above. The sequences for all markers/genes were available on Genbank/ EMBL.

Forward and reverse primers are given in the table A hereabove.

PCR condition were as followed

For each reaction in 30 microliters : DNA :60 ng; Taq DNA polymerase (Amersham) : 0.9 unit; Buffer 10x : 3 microliter; dNTP's : 0.2 mM each; MgCl2 : 1.5 mM; BSA 0.8mg/ ml; primers 1.5 ng/microliter each; glycerol 5%.

Polymerisation was done in a perkin Elmer 9600 : 1' at 95°C, followed by 35 cycles of (30" at 94°C, 30" at 60°C, 1'30" at 72°C) followed by 1'30" at 72°C.

The sequencing of 186 maize amplicon was then done with the primers used for DNA allele amplification. DNA sequences were edited and aligned. Sequence surronding polymorphism (see table I here-under were collected from these alignments.

LEGEND OF TABLE I (with references to the Bt2 gene for instance.)

Column 1 (Bt2) represents the name of the marker or gene.

Column 2 (Bt2-G2/G6-1) represents :

- the name of the maker (Bt2)

- the genotype number (G2)
- the second genotype number (G6)
- and the number of the SNP (single nucleotide polymorphism). So, in this case, it is a SNP found on a sequence nucleotide Bt2 between the genotypes (strains of maize) G2 and G6 and this SNP was numbered 1 (Sometimes there are several SNP between two genotypes for the same sequence)

Column 3 represents : similar to column 2, but with the codification of the marker/gene.

Column 4 represents sequence holding the SNP. Into brackets : [G/T] means that the sequence of G2, at this position of Bt2 gene, is G, while for G6, it is T.

On the other hand, /G (CSU61-G1/G5-1A) means deletion of the base pair G in G1 compared to G5.

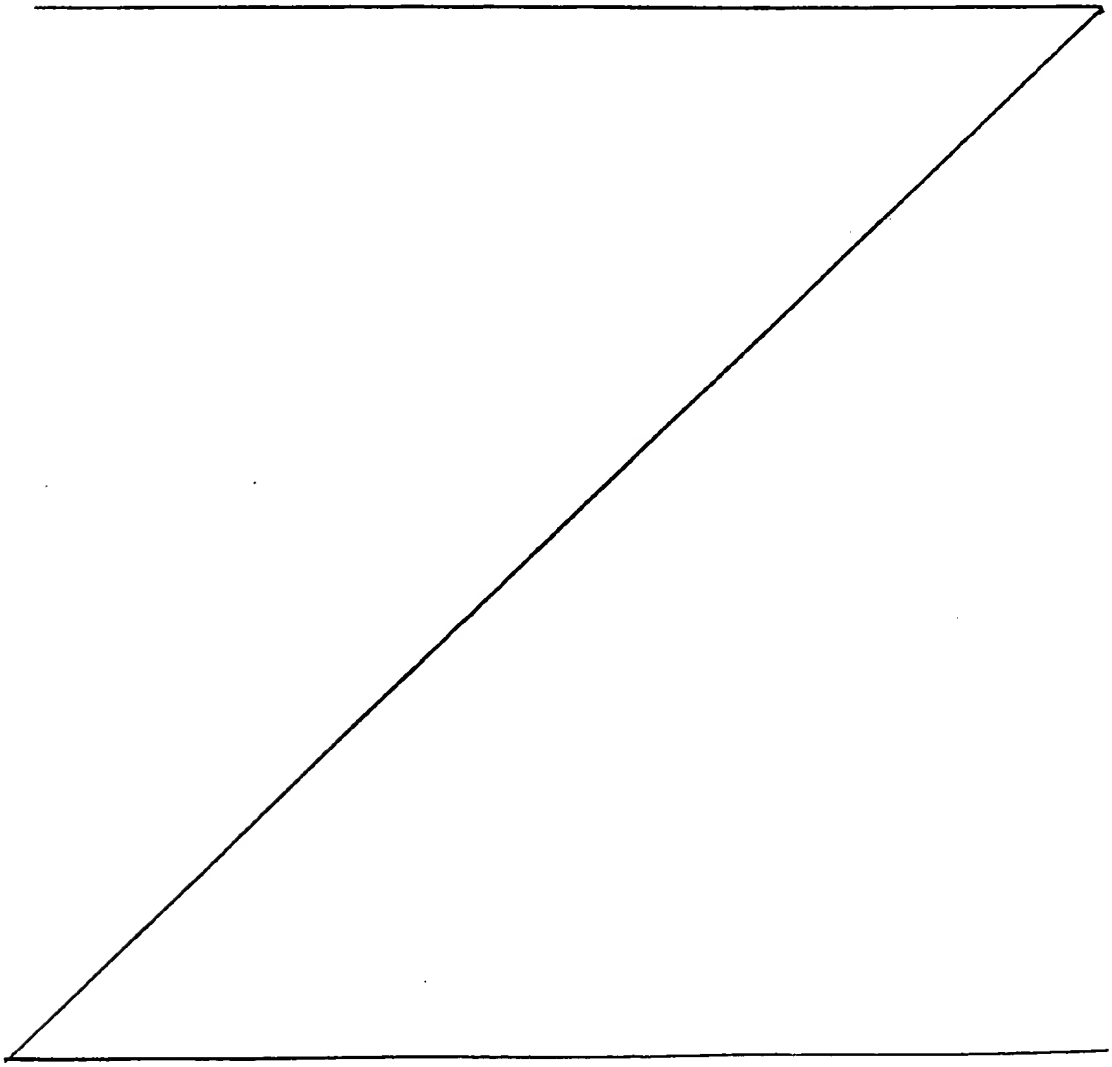

TABLE I

			csnp1d
Bt2	Bt2-G2/G6-1	S71G2/G6-1	ATAATACTTGATATGCCATT[G/T]TGTCCTCTTATTTTTAACAT
Ssu	Ssu- G1/G5-1	S67G1/G5-1	ATGGCCTCGTCGGCCACTGC[A/C]GTCGCTCCGTTCCATGGGCT
Ssu	Ssu-G1/G3-1	S67G1/G3-1	GCCGCTCCTCCAGAAGCCTC[G/A]GCAACGTCAGCAACGGCGGA
Ssu	Ssu-G1/G3-2	S67G1/G3-2	GTGTTGCCCATCCCATCCCA[A/T]TTCCCAACCCCAAACGAACC
Ssu	Ssu-G1/G3-3	S67G1/G3-3	GTACCTGCCGCCGCTGTCGA[CG/AC]GGACGACCTGCTGAAGCAGG
Bt1	Bt1-G2/G3-1	S66G2/G3-1	AGTGAGCCCGCTTCTTATTC[/T]TAAGGTGATAGGTTTCTAAA
Bt1	Bt1-G2/G3-2	S66G1/G3-1	AATGTAATGGTACTCCGCGC[T/C]ATGGCTCTGGTACTTAGGAA
Bt1	Bt1-G2/G3-3	S66G1/G2-1	AAATAGGCTCGGGCAATTAT[C/]CAGCTTAGGGACAGCAAGCG
Bre1	Bre1-G3/G6-1	S65G3/G6-1	TCCGCCCTGCCTCCGGTTTT[A/T]GCCCGACCTTCGAAACATTC
Bre1	Bre1-G3/G6-2	S65G3/G6-2	ACCACTGACGTAGCACCTCC[G/T]ACTTCTCGTTGTAAAACCCC
Bre1	Bre1-G3/G5-1	S65G3/G5-1	GGAGGTTCGCCTCATGTTAT[C/T]GTTGACGAGCCACATCCACT
Bre1	Bre1-G4/G6-1	S65G4/G6-1	GCTCCGACTTCCAATCTTGA[A/C]CCTCCACCCTGCCTCCGGTT
ASG12	ASG12-G1/G3-1	S64G1/G3-1	CTGGTTGAAATGTGTTGAAG[C/A]TACTAGTGATGAACTGCTTG
ASG12	ASG12-G1/g3-2A	S64G1/g3-2A	GCTGCTCCAAGCGAGCCCGC[C/G]CCGAAAAAGGAAAAAGGTGA
ASG12	ASG12-G1/g3-2B	S64G1/g3-2B	GCTGCTCCAAGCGAGCCCGC[C/G]CCGAAAAAGGAAAAAGTTGA
ASG12	ASG12-G1/g3-3A	S64G1/g3-3A	CGCCCCGAAAAAGGAAAAAG[G/T]TGAAGGTCCTTACTCACCGA
ASG12	ASG12-G1/g3-3B	S64G1/g3-3B	CGCGCCGAAAAAGGAAAAAG[G/T]TGAAGGTCCTTACTCACCGA
ASG12	ASG12-G1/gG3-4A	S64G1/gG3-4A	GAACCGGCCACAGTGCCTGA[T/A]TTTGGCGGTGAGACCTCTTC
ASG12	ASG12-G1/g3-4A	S64G1/g3-4A	GAACCGGCCACAGTGCCTGA[T/A]TTTGGCGGTGAGACTTCTTC
Sh2	Sh2-G5/G6-1	S63G5/G6-1	CAATTGTTACCTGAGCAAGA[T/]TTTTGTGTACTTGACTTGTT
Sh2	Sh2-G4/G6-1	S63G4/G6-1	TACTGAGAGAATGCAACATC[C/G]AGCATTCTGTGATTGGAGTC
Sh2	Sh2-G4/G5-1A	S63G4/G5-1A	TTTTAGTGTACTTGACTTGT[C/T]CTCCTCCACAGATGAAATAT
Sh2	Sh2-G4/G5-1B	S63G4/G5-1B	TTTTTGTGTACTTGACTTGT[C/T]CTCCTCCACAGATGAAATAT
Sh2	Sh2-G3/G6-1	S63G3/G6-1	TCTGTGATTGGAGTCTGCTC[G/A]CGTGTCAGCTCTGGATGTGA
Sh1	Sh1-G5/G6-1	S57G5/G6-1	AACTACAAAAAGCATCTCCT[G/T]GGATTTGGCTATCTCCTTTT
Sh1	Sh1-G2/G5-1	S57G2/G5-1	TTAGCGCGAAAAAAAAACTC[/T]TTTTTTTTTGTCCTTTTACT
Sh1	Sh1-G2/G3-1	S57G2/G3-1	TCAATCCAATCAATTTAATTT[T/C]CTTCCTTTAAAAATATTATC
Sh1	Sh1-G1/G2-1	S57G1/G2-1	TTACTACGAAAAACTCTTGA[G/T]TCTAGGAATTTGAATTTGTG
Sh1	Sh1-G1/G2-2A	S57G1/G2-2A	CTTCTTGGATTTTGCTATCT[T/C]CTTTTACTACGAAAAACTCT
Sh1	Sh1-G1/G2-2B	S57G1/G2-2B	CTCCTTGGATTTTGCTATCT[T/C]CTTTTACTACGAAAAACTCT
Sh1	Sh1-G1/G2-3A	S57G1/G2-3A	TTTTACTACGAAAAGCATCT[T/C]CTTGGATTTTGCTATCTTCT
Sh1	Sh1-G1/G2-3B	S57G1/G2-3B	TTTTACTACGAAAAGCATCT[T/C]CTTGGATTTTGCTATCTCCT
Sh1	S57G1/G2-4	S57G1/G2-4	GAAGCCAAATCCTATTATTT[T/C]CTGCCTCTAGGGTCTGAATG
UAZ77	UAZ77-G4/G6-1	S56G4/G6-1	GTACACTGTTACAATCACAC[T/G]TAGTGAAGCGCAACACAGAT
UAZ77	UAZ77-G4/G6-2	S56G4/G6-2	GCCTTATCATCCTCTAGGTA[T/A]TGGAGACGAGTGACCAGTCT
UAZ77	UAZ77-G4/G6-3	S56G4/G6-3	CTTTTCTTCAGACCCGAGCC[C/T]CCAATCGCGCCCTTCTGTGC
UAZ77	UAZ77-G4/G6-3	S56G4/G6-3	CTTTTCTTCAGACCCGAGCC[C/T]CCAATCGCGCCCTTTTGTGC
UAZ77	UAZ77-G4/G5-1A	S56G4/G5-1A	GAGCCCCCAATCGCGCCCTT[C/T]TGTGCCTTGGCCTTGAGCTC
UAZ77	UAZ77-G4/G5-1A	S56G4/G5-1A	GAGCCTCCAATCGCGCCCTT[C/T]TGTGCCTTGGCCTTGAGCTC
UAZ171	UAZ171G1/G3-1	S55G1/G3-1	GAAGGAGCAGCAGCGCAAGG[A/]ACGTGTTCCAAGTCAACGTC
UMC17	UMC117-G2/G3-1	S54G2/G3-1	GTAGAAAGTTAGCAAAAACA[T/]TTTTTTAGTGAAAAAACATA
UMC17	UMC117-G2/G3-2	S54G2/G3-2	ATTGTGGCTAGAAACTTTGG[/T]TTTTTTTAAATTATGGTCAT
CSU109	CSU109-G5/G6-1	S53G5/G6-1	GCAAACCAACACCAATCTTC[G/C]AAATGAGCAAAGCAGAGACT
CSU109	CSU109-G5/G6-2	S53G5/G6-2	CAGATCGGTTGTCCTCAGAG[A/]AAGTCACCTACCTGCAAACC
CSU109	CSU109-G5/G6-3	S53G5/G6-3	AATTCTACATAGGAGTCATG[C/T]ACAAGTACTTGTTTAAAGGA
CSU109	CSU109-G5/G6-4	S53G5/G6-4	ACAAGTACTTGTTTAAAGGA[C/]CATGCCGGAATACACGCTGC
CSU109	CSU109-G5/G6-5A	S53G5/G6-5A	GAGCGAGATCGATCCTGTTG[T/C]CATCCATCACTGCCATAGGA
CSU109	CSU109-G5/G6-5B	S53G5/G6-5B	GAGCGAGATCGATCCTGTTG[T/C]CATCCATCACTGCCGTAGGA
CSU109	CSU109-G4/G6-1	S53G4/G6-1	TAGTCATAGCAACAGCATGC[G/A]TCGTGATGTAGCGTTCACCC
CSU109	CSU109-G4/G6-2	S53G4/G6-2	CAATTGAAGAGGAAAAAAAA[/T]TCTACATAGGAGTCATGTAC
CSU109	CSU109-G4/G5-1	S53G4/G5-1	CAGAGACTCCACAAGGCGAA[A/C]GGAGTCCACAATAGTTCGTC
CSU109	CSU109-G3/G5-1	S53G3/G5-1	CCCACGGCGGGAGATGGTGG[T/]TAGAAGCGGAACCACCGAGC
CSU109	CSU109-G2/G6-1	S53G2/G6-1	ACTTGTTTAAAGGACATGCC[G/]GGAATACACGCTGCCCAGGC
CSU109	CSU109-G2/G3-1	S53G2/G3-1	CCCAGGCCTTCCCACGGCGG[A/G]GATGGTGGTTAGAAGCGGAA
CSU109	CSU109-G1/G6-1	S53G1/G6-1	CAAAGCAGAGACTCCACAAG[A/G]CGAACAGAGTCCGCAATAGT
CSU109	CSU109-G1/G6-2	S53G1/G6-2	GAACAGAGTCCGCAATAGTT[T/C]ATCCTAATGCTACTTCGAGC
UMC130	UMC130-G3/G6-1	S48G3/G6-1	GATTCAGAAACAGTGGCGGC[A/G]GATGTAGCATCAACACGCCC
CSU61	CSU61-G5/G6-1	S45G5/G6-1	ATGAGTATATTCAAGTCATA[T/C]TGTGAACTAGAATGTTATTT
CSU61	CSU61-G5/G6-2A	S45G5/G6-2A	CCTAGACGCTGACCGCCACA[G/A]ACGGCGGCGGCTGCCAAATC
CSU61	CSU61-G5/G6-2B	S45G5/G6-2B	CCTAAACGCTGACCGCCACA[G/A]ACGGCGGCGGCTGCCAAATC
CSU61	CSU61-G5/G6-3	S45G5/G6-3	TGAACAAACCATGCGCTACC[C/T]AGCTAGGTGTTTTAAAGTAA
CSU61	CSU61-G4/G6-1	S45G4/G6-1	TCCGCGGAAACAACATCCGA[G/T]TTCTTGAGGATAACCCAGCT
CSU61	CSU61-G4/G5-1	S45G4/G5-1	GGGAGGGGAAAAAAAAAGAA[G/A]AGCGTTGGTTGCGGTTCAGT
CSU61	CSU61-G4/G5-2	S45G4/G5-2	GGCGGCTGCCAAATCCGCGG[/A]AAACGACATCCGAGTTCTTG
CSU61	CSU61-G2/G4-1A	S45G2/G4-1A	CTAGAATGTTATTTCTTCAC[C/A]GTTGACCATGGAAAAAAACA
CSU61	CSU61-G2/G4-1B	S45G2/G4-1B	CTAGAATGTTATTTCTTCAC[C/A]GTTGACCATGGAAAGAAACA
CSU61	CSU61-G2/G4-2A	S45G2/G4-2A	TTCACCGTTGACCATGGAAA[A/G]AAACAGTAATAAGTTCTTGT
CSU61	CSU61-G2/G4-2B	S45G2/G4-2B	TTCACAGTTGACCATGGAAA[A/G]AAACAGTAATAAGTTCTTGT
CSU61	CSU61-G1/G6-1	S45G1/G6-1	TTCTTCACAGTTGACCATGG[/A]AAAAAAACAGTAATAAGTTC
CSU61	CSU61-G1/G5-1A	S45G1/G5-1A	GAACCCACCGTGCCCTGGGA[/G]GGGAAAAAAAAAGAAGAGCG
CSU61	CSU61-G1/G5-1B	S45G1/G5-1B	GAACCCACCGTGCCCTGGGA[/G]GGGAAAAAAAAAGAAAGCG
CSU61	CSU61-G1/G5-2A	S45G1/G5-2A	TGGGAGGGAAAAAAAAAGAA[G/A]AGCGTTGGTTGCGGTTCAGT
CSU61	CSU61-G1/G5-3	S45G1/G5-3	CGTACCAGCTAGGAATCGTA[A/G]AAAGCCTAGACGCTGACCG
UMC95	UMC95-G5/G6-1	S44G5/G6-1	GCTGCGTCAATCATCACTTC[T/A]CCCACAGGCGTCAAGTACAG
UMC95	UMC95-G3/G4-1	S44G3/G4-1	GACAGATTCCAAAGTAGTCG[C/T]CGGCCAGGTCGAAAAAGAAT
UMC95	UMC95-G2/G6-1	S44G2/G6-1	GGCGCTGCGTCAATCATCAC[A/T]TCACCCACAGGCGTCAAGTA

UMC95	UMC95-G2/G4-1A	S44G2/G4-1A	TCGGTGTCACCACATGCATA[T/G]TCAGGACAGATTCCAAACTA
UMC95	UMC95-G2/G4-1B	S44G2/G4-1B	TCGGTGTCACCACATGCATA[T/G]TCAGGACAGATTCCAAAGTA
UMC95	UMC95-G2/G4-2A	S44G2/G4-2A	GTCGCCGGCCAGGTCGAAAA[G/A]GAATACTCAGCAAAAGACCC
UMC95	UMC95-G2/G4-2B	S44G2/G4-2B	GTCGTCGGCCAGGTCGAAAA[G/A]GAATACTCAGCAAAAGACCC
UMC95	UMC95-G2/G3-1A	S44G2/G3-1A	TATTCAGGACAGATTCCAAA[C/G]TAGTCGCCGGCCAGGTCGAA
UMC95	UMC95-G2/G3-1B	S44G2/G3-1B	TAGTCAGGACAGATTCCAAA[C/G]TAGTCGCCGGCCAGGTCGAA
UMC95	UMC95-G1/G6-1	S44G1/G6-1	GCGTCAAGTACAGATACGCA[A/G]CACGCCTCAGCTTCGCCTTG
UMC95	UMC95-G1/G2-1	S44G1/G2-1	CCTGGGACTCCGCAAATTGC[G/A]AGCACTCGGTGTCACCACAT
Wx1	Wx1-G2/G6-1	S43G2/G6-1	GCTGGTTCATTATCTGACCT[G/T]GATTGCATTGCAGCTACAAG
Wx1	Wx1-G2/G6-2	S43G2/G6-2	CTGGATTGCATTGCAGCTAC[A/G]AGAAGCCCGTGGAAGGCCGG
Wx1	Wx1-G2/G6-1B	S43G2/G6-1B	GCTGGTTCATTATCTGACCT[G/T]GATTGCATTGCAGCTACGAG
Wx1	Wx1-G2/G6-2B	S43G2/G6-2B	CTTGATTGCATTGCAGCTAC[A/G]AGAAGCCCGTGGAAGGCCGG
Wx1	Wx1-G2/G6-3	S43G2/G6-3	TCAGCCCCTACTACGCCGAA[G/]GAGCTCATCTCCGGCATCGC
Wx1	Wx1-G2/G5-1	S43G2/G5-1	TACCCGGAGCTGAACCTCCC[C/G]GAGAGATTCAAGTCGTCCTT
Wx1	Wx1-G2/G4-1	S43G2/G4-1	TGCATGTGAACATTCATGAA[T/C]GGTAACCCACAACTGTTCGC
Wx1	Wx1-G6/G1-1	S43G6/G1-1	CTCCTACCAGGGCCGGTTCG[T/]CCTTCTCCGACTACCCGGAG
Wx1	Wx1-G1/G6-1	S43G1/G6-1	TGAATGGTAACCCACAACTG[C/T]TCGCGTCCTGCTGGTTCATT
Wx1	Wx1-G1/G5-1	S43G1/G5-1	GCCGACAGGGTCCTCACCGT[G/C]AGCCCCTACTACGCCGAAGA
Wx1	Wx1-G2/G6-1	S43G2/G6-1	GCTGGTTCATTATCTGACCT[G/T]GATTGCATTGCAGCTACAAG
Wx1	Wx1-G2/G6-1B	S43G2/G6-1B	GCTGGTTCATTATCTGACCT[G/T]GATTGCATTGCAGCTACGAG
Wx1	Wx1-G2/G6-2	S43G2/G6-2	CTGGATTGCATTGCAGCTAC[A/G]AGAAGCCCGTGGAAGGCCGG
Wx1	Wx1-G2/G6-2B	S43G2/G6-2B	CTTGATTGCATTGCAGCTAC[A/G]AGAAGCCCGTGGAAGGCCGG
Wx1	Wx1-G2/G6-5	S43G2/G6-5	TCAGCCCCTACTACGCCGAA[G/]GAGCTCATCTCCGGCATCGC
Wx1	Wx1-G2/G4-1	S43G2/G4-1	TGCATGTGAACATTCATGAA[T/C]GGTAACCCACAACTGTTCGC
Wx1	Wx1-G2/G3-1	S43G2/G3-1	CTGGTGGTGGTGCTTCTCTG[AAAC/]TGAAACTGAAACTGACTGCA
Wx1	Wx1-G2/G3-3	S43G2/G3-3	GACCATCTTCACGTACTACC[TACC/]AGACCGCTTTCTGCATCCAC
Wx1	Wx1-G1/G6-1	S43G1/G6-1	CTGACCATCTTCACGTACTA[CCTA/]CCAGACCGCTTTCTGCATCC
Wx1	Wx1-G6/G1-1	S43G6/G1-1	CTCCTACCAGGGCCGGTTCG[T/]CCTTCTCCGACTACCCGGAG
Wx1	Wx1-G6/G1-1	S43G6/G1-1	GAGATTCAAGTCGTCCTTCG[G/]ATTTCATCGACGGGTCTGTT
Wx1	Wx1-G1/G6-2	S43G1/G6-2	TGAATGGTAACCCACAACTG[C/T]TCGCGTCCTGCTGGTTCATT
Wx1	Wx1-G1/G6-3	S43G1/G6-3	GCCGACAGGGTCCTCACCGT[G/C]AGCCCCTACTACGCCGAAGA
Wx1	Wx1-G1/G5-1	S43G1/G5-1	TCTGACCATCTTCACGTACT[ACCT/]ACCAGACCGCTTTCTGCATC
Wx1	Wx1-G1/G4-1	S43G1/G4-1	CTTGATTGCATTGCAGCTAC[G/A]AGAAGCCCGTGGAAGGCCGG
Wx1	Wx1-G1/G4-1	S43G1/G4-1	CTGGATTGCATTGCAGCTAC[G/A]AGAAGCCCGTGGAAGGCCGG
Wx1	Wx1-G1/G3-1	S43G1/G3-1	GCTGGTTCATTATCTGACCT[T/G]GATTGCATTGCAGCTACGAG
Wx1	Wx1-G5/G6-1	S43G5/G6-1	AGAGATTCAAGTCGTCCTTC[G/]GATTTCATCGACGGGTCTGT
UMC109	UMC109-G2/G6-1	S42G2/G6-1	CTCCATGAAAAAGGTGCCGCI/G]TACTCTCTCAGTCAGCTACT
UMC109	UMC109-G2/G3-1A	S42G2/G3-1A	CTGCACTCCGATTGAGGGTC[C/G]GAAGCAGGGCAGCGCGTGTG
UMC109	UMC109-G2/G3-1B	S42G2/G3-1B	CTGCACTCCGATTGAGGGTC[C/G]GAAGCAGGGCAGCGCGTTGT
UMC109	UMC109-G2/G3-1C	S42G2/G3-1C	CTGCACTCCGATTGAGGGTC[C/G]GAAGCAGGGCAGCGCGTTTG
UMC109	UMC109-G2/G3-1D	S42G2/G3-1D	CTGCACTCCGATTGAGGGTC[C/G]GAAGCAGGGCAGCGCGTTTT
UMC80	UMC80-G3/G5-1	S34G3/G5-1	CATGCCTCTGTTGATATTTT[G/C]GTGCACCTTTTGCTTGCAAC
UMC80	UMC80-G3/G5-2	S34G3/G5-2	GATTTTGTAGGTTGATGCAT[C/T]GTTTGATCTTTCTTATCTCC
UMC80	UMC80-G3/G5-3	S34G3/G5-3	TGCTTGCAACTAAATTAATC[A/G]TGCTCTATTTGACTAAGAGT
UMC80	UMC80-G3/G4-1	S34G3/G4-1	ACATGTCCAGGACGCATGGT[C/]CCCAATATTGTTGTTGGAAG
UMC80	UMC80-G3/G4-2	S34G3/G4-2	TTGATCTTTCTTATCTCCTT[/C]CGAATTTGTTCTGTGTTATA
UMC80	UMC80-G3/G4-2B	S34G3/G4-2B	TTGATCTTTCTTATCTCCTT[/C]CGAATTTGTTCTGTGTATAC
UMC80	UMC80-G2/G5-1	S34G2/G5-1	TGTAGGACTTGGAGAGCTTG[A/G]TAATTTACACATGCCTCTGT
UMC80	UMC80-G2/G5-2	S34G2/G5-2	CATGCCTCTGTTGATATTTT[G/C]GTGCACCTTTTGCTTGCAAC
UMC80	UMC80-G2/G5-3	S34G2/G5-3	GATTTTGTAGGTTGATGCAT[C/T]GTTTGATCTTTCTTATCTCC
UMC80	UMC80-G2/G3-1	S34G2/G3-1	GAGACATTTCCTACTCAATA[C/T]AATTATTTGATGAAATTATT
UMC254	UMC254-G5/G6-1A	S33G5/G6-1A	AGTATCACAGACTAATCTGA[A/G]TATCTGGTTGCCACGAAAAC
UMC254	UMC254-G5/G6-1B	S33G5/G6-1B	AGTATCACAGACTAATCTGA[A/G]TATCTGGTTGCCACAAAAAC
UMC254	UMC254-G5/G6-2	S33G5/G6-2	TCAAAGTGGTGCAATCGCAA[T/C]CCACTTGGGCTTGCCGTGGT
UMC254	UMC254-G5/G6-3	S33G5/G6-3	CCACTTGGGCTTGCCGTGGT[C/]CGTATCGTACGCAGGTAGCA
UMC254	UMC254-G5/G6-4	S33G5/G6-4	AGCATTTTTTGTTTTGTTTTT[T/C]CCTTGGCAGACAACAGACAG
UMC254	UMC254-G5/G6-5A	S33G5/G6-5A	CAGTCCCGAGAATCCCAAAT[C/]CAGAAAAAGGTTTTGTTTTT
UMC254	UMC254-G5/G6-5B	S33G5/G6-5B	CAGTCCCGAGAATCCCAAAT[C/]CAGAAAAAGGTTTTGTTTTA
UMC254	UMC254-G4R/G6-1A	S33G4R/G6-1A	GGCAGACAACAGACAGATCA[AG/CA]CATGCTTGCATTTACTCCCA
UMC254	UMC254-G4R/G6-1B	S33G4R/G6-1B	GGCAGACAACAGACAGATCA[AG/CA]CATGCTTGCATTTACTCTCA
UMC254	UMC254-G3R/G6-1A	S33G3R/G6-1A	GTGATCACAGACTAATCTGA[A/G]TATCTGGTTGCCACGAAAAC
UMC254	UMC254-G3R/G6-1B	S33G3R/G6-1B	GTGATCACAGACTAATCTGA[A/G]TATCTGGTTGCCACAAAAAC
UMC254	UMC254-G3R/G6-2A	S33G3R/G6-2A	TCTGAATATCTGGTTGCCAC[G/A]AAAACCGGGACACAAGAGAG
UMC254	UMC254-G3R/G6-2B	S33G3R/G6-2B	TCTGAGTATCTGGTTGCCAC[G/A]AAAACCGGGACACAAGAGAG
UMC254	UMC254-G3/G6-3	S33G3/G6-3	TCAGTCAAACTCAGTCCCGA[A/G]AATCCCAAATCAGAAAAAGG
UMC254	UMC254-G3/G5-1A	S33G3/G5-1A	GGTTGCCACGAAAACCGGGA[C/G]ACAAGAGAGAAACTCAGAGT
UMC254	UMC254-G3/G5-1B	S33G3/G5-1B	GGTTGCCACGAAAACCGGGA[C/G]ACAAGAGAGAAACTCAAAGT
UMC254	UMC254-G2R/G3-1A	S33G2R/G3-1A	ACGCATGCTTGCATTTACTC[C/T]CAGTCAAACTCAGTCCCGAA
UMC254	UMC254-G2R/G3-1B	S33G2R/G3-1B	ACACATGCTTGCATTTACTC[C/T]CAGTCAAACTCAGTCCCGAA
UMC254	UMC254-G1R/G2-1	S33G1R/G2-1	TATTATTCAATTTTGAATAA[/G]GAAGGAAATTTTAGCACCTC
ASG49	ASG49-G3/G5-1	S32G3/G5-1	ATTAATAAATGCATCCTCTG[C/G]TAAAAAAACCCATTTTGAAT
ASG49	ASG49-G3/G5-2	S32G3/G5-2	ATGAATTGAAGCTCTGAATA[C/T]AGAATCCACCATTCTTCCGA
ASG49	ASG49-G3/G5-3	S32G3/G5-3	GAATCCACCATTCTTCCGAA[A/G]CTGCTTCCTACAAAACTCGA
ASG49	ASG49-G3/G5-4	S32G3/G5-4	GAAAGGATGTGTTTTTGATA[G/A]CCTTCAGTCTTTCAGATGGA
ASG8	ASG8-G3/G5-1	S31G3/G5-1	CAATGTCTTGTTCGTTATCA[A/G]CGAAAGTTTGAATCCCCACA
ASG8	ASG8-G3/G4-1	S31G3/G4-1	TGTATCGGCTAGTCTGGATG[G/A]TCGCACTGGCACTCAGTGCT

csnp1d

UMC132	UMC132-G4/G5-1	S29G4/G5-1	TCTATTCAGCAGTCTGAGAA[GCA/CT]AGGATGGTCGGCTTCTTCAG
UMC132	UMC132-G1/G5-1	S29G1/G5-1	CCTTACACTATTAACAGGCC[C/T]GTGATCTACCTGAATGCCTG
UMC21	UMC21-G5/G6-1	S28G5/G6-1	CAAGAAGCCTCTTCAGTGTC[A/C]GTCGTAGCTTCCTCAAGACC
UMC21	UMC21-G5/G6-2	S28G5/G6-2	AGACCTTCCTGATGTGCGGA[T/C]GCTAATCCATGGAGCAGGGA
UMC21	UMC21-G5/G6-2B	S28G5/G6-2B	AAGACCTCCTGATGTGCGGA[T/C]GCTAATCCATGGAGCAGGGA
UMC21	UMC21-G5/G6-3	S28G5/G6-3	CTAATCCATGGAGCAGGGAG[G/A]AAGGGGCGAGGGGCAGCAAG
UMC21	UMC21-G4/G5-1	S28G4/G5-1	TCGTCGCGAATACAGCCGGG[G/C]GAGGGGGTGGTCGCGACTGG
UMC21	UMC21-G3/G6-1	S28G3/G6-1	GTCGTAGCTTCCTCAAGACC[T/]TCCTGATGTGCGGACGCTAA
UMC21	UMC21-G3/G4-1	S28G3/G4-1	GAGTCGTCGCGAATACAGCC[A/G]GGGGAGGGGGTGGTCGCGAC
UMC21	UMC21-G3/G4-2	S28G3/G4-2	AGGGGGTGGTCGCGACTGGA[T/G]CGCCCGAGCAGCGAGCAAGC
UMC21	UMC21-G3/G4-3	S28G3/G4-3	AAGCACATGTTTTAACCTTT[T/G]ATTCAAACTTTCCAGCCGTT
UMC21	UMC21-G3/G4-3B	S28G3/G4-3B	AAGCACATGTTTTAACCTTT[T/G]ATTCAAACTTTCCAGCGTTA
UMC21	UMC21-G2/G6-1	S28G2/G6-1	GAATGTTGCTGTTATATTAC[T/C]CGTAGGTGACAAAGGGTTCA
UMC21	UMC21-G2/G4-1	S28G2/G4-1	AGAAAAATTTACATAAAAAA[G/C]CACACTCCATGATTGTTAAA
UMC21	UMC21-G2/G4-1B	S28G2/G4-1B	AGAAAAATTTACATAAAAAA[G/C]CACACTCCATGATTGTTTAA
UMC21	UMC21-G2/G3-1	S28G2/G3-1	CTTTTATTCAAACTTTCCAG[/C]CGTTAATTTGTTATCCGTTG
UMC21	UMC21-G6/G1-1	S28G6/G1-1	TGTTGAACATGCTCTCAGGA[/CC]CCCCCTATTGTGACACAGCA
UMC21	UMC21-G1/G3-1	S28G1/G3-1	TACATCTTAACAAGCACATG[TG/TTT]TAACCTTTTATTCAAACTTT
UMC65	UMC65-G3/G6-1A	S27G3/G6-1A	AGTAATGTGTGACTGTGGGC[C/G]CGTGTGACAGCTTTTACGTA
UMC65	UMC65-G3/G6-1B	S27G3/G6-1B	AGTAGTGTGTGACTGTGGGC[C/G]CGTGTGACAGCTTTTACGTA
UMC65	UMC65-G3/G6-2	S27G3/G6-2	TTCGCTTGGTAGCCGTAGCA[G/A]TATACTTTTACCGGCCACAG
UMC65	UMC65-G3/G6-3	S27G3/G6-3	GGGCTTTGGGTTGTGAACTT[CCA/C]AAAAAAAAAAAAAATTTCCC
UMC59	UMC59-G5/G6-1	UMC59-G5/G6-1	CCAAGAAAGATTAATGCTGG[/T]TAAAATATTGTTTCCAGTCT
UMC59	UMC59-G5/G6-2	UMC59-G5/G6-2	AAAATCAGGACTGCGAAAAA[A/C]CCAAGAAAGATTAATGCTGG
UMC59	UMC59-G5/G6-2B	UMC59-G5/G6-2B	AAAATCAGGACTGCGAAAAA[A/C]CCAAGAAGATTAATGCTGGT
UMC59	UMC59-G5/G6-3	UMC59-G5/G6-3	AAAGTGTGTGTTGTTGCCCA[G/A]ATGATTCCATTCCACACAAG
UMC59	UMC59-G4/G5-1	UMC59-G4/G5-1	AGGACTGCGAAAAAACCAAG[/A]AAGATTAATGCTGGTAAAAT
UMC59	UMC59-G4/G5-2	UMC59-G4/G5-2	ATGCTGGTAAAATATTGTTT[/C]CAGTCTTTCACAAAGTGTGT
UMC59	UMC59-G3/G4-1	UMC59-G3/G4-1	CTACAAAAATCAGGACTGCG[/A]AAAAACCAAGAAGATTAATG
UMC59	UMC59-G3/G4-2	UMC59-G3/G4-2	TTGTTTCAGTCTTTCACAAA[/GT]GTGTGTGTGCCAGATGATTC
UMC59	UMC59-G3/G4-3	UMC59-G3/G4-3	TCACACACCGACCTGCCTGG[/T]TATCAGGAACCATCCTCCTG
Ae1	Ae1-G4/G5-1	S23G4/G5-1	GGTGAATTGGTGATGCATGC[T/G]GGGGGTGCTCGAGTTGGATG
Ae1	Ae1-G4/G5-2	S23G4/G5-2	TTCCAGTCGGATGAACTGGA[T/G]GTTCGTCATCCACTCGTCAC
Ae1	Ae1-G3/G6-1	S23G3/G6-1	GGTGAATTGGTGATGCATGC[A/T]GGGGGTGCTCGAGTTGGATG
Ae1	Ae1-G5/G3-1	S23G5/G3-1	TTAAGTGAAGATGCCCAAAC[C/G]GTTAAACTTTCCATGGAACT
Ae1	Ae1-G5/G3-1B	S23G5/G3-1B	ATTAATGAAGATGCCCAAAC[C/G]GTTAAACTTTCCATGGAACT
Ae1	Ae1-G1/G6-1	S23G1/G6-1	TGATTCGGGTCTGTATGCGA[G/T]TGTTGTGGTGGTGAACTGGT
Ae1	Ae1-G1/G5-1	S23G1/G5-1	CGGGTCTGTATGCGAGTGTT[G/A]TGGTGGTGAACTGGTGAATT
Ae1	Ae1-G1/G4-1	S23G1/G4-1	GTTCGCGGTTTCTGGGGCCG[G/T]GGGCGGTGCTCGGTGGGGCC
UMC90	UMC90-G5/G6-1	S22G5/G6-1	CAGATTGGTGTCGTTTACTA[A/G]AATTCAGTTCTGTCCATTTG
UMC90	UMC90-G5/G6-2	S22G5/G6-2	AAGTAAGCATTCTTTATATG[/T]TACTTCCCATGATAAACTTT
UMC90	UMC90-G5/G6-3	S22G5/G6-3	CAAAGGGCTTACTGTACTTT[/C]CATCTTATTGGCAGGGCACC
UMC66	UMC66-G5/G6-1	S19G5/G6-1	ACTTGGCCGGGGACGTCGAC[G/A]ATCGTCGTAGCACTACTGGT
UMC66	UMC66-G5/G6-2	S19G5/G6-2	AGTACATGGCGAGCGTTGTA[G/C]CAGCTGCTTAGGTGATGTGG
Adh2	Adh2-G4/G6-1	S17G4/G6-1	CTATTTCCAAGCTAACAACC[C/G]CTCTTGGTCCCAACATCCTG
Adh2	Adh2-G3/G6-1	S17G3/G6-1	GGTTCTAAACATAGCTCGTC[C/A]ATTCATGATTCATCTCGAGC
UMC63	UMC63-G4/G6-1	S16G4/G6-1	TCAGCAAGCCTCCAAGGCTC[C/A]AATGGTCCAGTTACTTGGTT
UMC63	UMC63-G2/G6-1	S16G2/G6-1	GTGTGTAGCTTCATTCGCAA[TG/AT]TTTGAACAGCCTCTGCAAGT
UMC63	UMC63-G2/G6-2A	S16G2/G6-2A	GTGCTTTCGTAAACCTAGAG[T/C]TGACCAGCTGTGATTTCGGT
UMC63	UMC63-G2/G6-2B	S16G2/G6-2B	GTGCTTTCGTAAACCTAGAG[T/C]TGACCAGCTGTGATTTCGAT
UMC63	UMC63-G2/G6-3A	S16G2/G6-3A	GCTGACCAGCTGTGATTTCG[G/A]TGTATTCCACGACCACGAGT
UMC63	UMC63-G1/G6-1	S16G1/G6-1	TGTGTAGCTTCATTCGCAAA[G/T]TTTGAACAGCCTCTGCAAGT
UMC63	UMC63-G1/G3-2A	S16G1/G3-2A	GTGCTTCCGTAAACCTAGAG[T/C]TGACCAGCTGTGATTTCGAT
UMC63	UMC63-G1/G3-2B	S16G1/G3-2B	GTGCTTCCGTAAACCTAGAG[T/C]TGACCAGCTGTGATTTCGGT
UMC63	UMC63-G1/G2-1	S16G1/G2-1	GTGTGTAGCTTCATTCGCAA[A/T]GTTTGAACAGCCTCTGCAAG
UMC102	UMC102-G5/G6-1	S14G5/G6-1	GCTCAGCTGCCGGAGTACGT[A/T]GGCTTGCTCTCCGGCCGGCC
UMC102	UMC102-G5/G6-1B	S14G5/G6-1B	ATAGCTCTGCCGGAGTACGT[A/T]GGCTTGCTCTCCGGCCGGCC
ASG24	ASG24-G5/G6-1	S13G5/G6-1	TTTCACAACTCAACTGATTG[A/T]CTTGCTTTGATGTGGATTCT
ASG24	ASG24-G2/G6-1	S13G2/G6-1	TTGGTAATTTCAGAGCTAGA[C/G]AACTTACTGTGGTACACGCC
UMC49	UMC49-G4/G6-1	S12G4/G6-1	ACCTTTGCTGTGTTTTTTTTT[T/G]GTATTCGAATGGAGGGAGTA
UMC49	UMC49-G2/G5-1	S12G2/G5-1	AAAACAGCCAAGGTGGTGGT[C/G]AAAGGAAGGTGTCAGAAGGT
UMC49	UMC49-G2/G5-2	S12G2/G5-2	TCTGTTCGTTCCATCTCTTT[A/G]CAGTAAATATCCGTAATTAC
UMC49	UMC49-G2/G5-3	S12G2/G5-3	CGTAATTACTTTGTTACTAC[TA/C]AGTAATTTTATATATATCCT
UMC49	UMC49-G2/G5-4	S12G2/G5-4	TATATATATCCTCATTTCAA[A/T]GAACAGTCAAAGTTAGTTTT
UMC49	UMC49-G2/G5-4B	S12G2/G5-4B	TATATATATCCTCATTTCAA[A/T]GAACAGTCAAAGTAGTTTTG
UMC49	UMC49-G2/G4-1	S12G2/G4-1	TATTTCTTATCCAGGATTGT[T/C]CTTTGGCCAAAGCATGGTAC
UMC49	UMC49-G2/G4-2	S12G2/G4-2	CGTTCCATCTCTTTACAGTA[A/G]ATATCCGTAATTACTTTGTT
UMC49	UMC49-G2/G4-3	S12G2/G4-3	ATCCGTAATTACTTTGTTAC[TA/AC]CTAAGTAATTTTATATATAT
UMC49	UMC49-G2/G3-4	S12G2/G3-4	GTAATTACTTTGTTACTACT[A/]AGTAATTTTATATATATCCT
UMC49	UMC49-G1/G6-1	S12G1/G6-1	CTGTGTTTTTTTTTGGTATT[G/C]GAATGGAGGGAGTATTATTT
UMC49	UMC49-G1/G6-1B	S12G1/G6-1B	GCTGTGTTTTTTTTGGTATT[G/C]GAATGGAGGGAGTATTATTT
UMC49	UMC49-G1/G5-1	S12G1/G5-1	ACTTAGATGATGACCAGGTG[A/]AGAGTTTGGCACCTTTGCTG
UMC49	UMC49-G1/G5-2	S12G1/G5-2	AGTTTGGCACCTTTGCTGTG[T/]TTTTTTTTGGTATTGGAATG
UMC49	UMC49-G1/G5-3	S12G1/G5-3	CTTTACTGATTGGGTTACAA[A/G]AGGTTATTTCTTATTCAGGC
UMC49	UMC49-G1/G5-4	S12G1/G5-4	AATTACTTTGTTACTACCAG[T/]TAATTTTATATATATCCTCC
UMC131	UMC131-G4/G6-1	S10G4/G6-1	AGCGACAGGGATGTCGAGCA[G/T]CTACGGAAGGCAATAATGAG

csnp1d

UMC131	UMC131-G4/G6-2	S10G4/G6-2	AATTTGGGAAAATCAATGCA[GAA/CAC]ATCAGTGATTAATCCACATA
UMC131	UMC131-G3/G6-1	S10G3/G6-1	GCATGGCGGAGTGAGGGAGG[TG/]TGTGTGTGTGTGGCTCCACA
UMC131	UMC131-G3/G6-2A	S10G3/G6-2A	GGCCGCTACGCCATTTAGCG[G/A]ATTTGGGAAAATCAATGCAG
UMC131	UMC131-G3/G6-2B	S10G3/G6-2B	GGCCGCTACGCCATTTAGCG[G/A]ATTTGGGAAAATCAATGCAC
UMC131	UMC131-G1/G6-1	S10G1/G6-1	CATCCCCGCCGGCAGAACAA[C/G]GTACGAGAAGGATGGAATGC
UMC53	UMC53-G5/G6-1	UMC53-G5/G6-1	GTCCCAGATCAGGTCCACGT[T/C]CGAGCTCGCTGTTCCCGCTT
UMC53	UMC53-G5/G6-2	UMC53-G5/G6-2	TGGTTCTTCACCACCACCGC[C/G]CCGGGCGCGCCCAGCGCCTC
UMC53	UMC53-G4/G6-1	UMC53-G4/G6-1	GCAGCCTCAGGTACACGGGG[/A]AAGTCGGAGTGGTTCTTCAC
UMC53	UMC53-G4/G6-2	UMC53-G4/G6-2	GCCGGGCGCGCCCAGCGCCT[/C]CGTCCCAGATCAGGTCCACG
UMC53	UMC53-G3/G6-1	UMC53-G3/G6-1	GCACGTCGTTGGTGAAGAAG[AC/CA]GCGGTACGGGTGCTTGTCGA
UMC53	UMC53-G3/G5-1	UMC53-G3/G5-1	AGGTACACGGGGAAGTCGGA[G/T]TGGTTCTTCACCACCACCGC
UMC53	UMC53-G3/G5-2	UMC53-G3/G5-2	CGACGGCGTCCAGCACCGAC[G/]CCTCCGCCTTCACCCCGCGC
UMC53	UMC53-G3/G4-1	UMC53-G3/G4-1	GTCCACGTCGAGCTCGCTGT[C/T]CCCGCTGCCCACGACGGCGT
UMC53	UMC53-G1/G4-1	UMC53-G1/G4-1	GCACGTCGTTGGTGAAGAAG[A/C]AGCGGTACGGGTGCTTGTCG
UMC161	UMC161-G2/G3-1	S06G2/G3-1	NAACCAAACCCTGACTATTA[T/C]AGGTAGATTAGACTAGACAC
UMC161	UMC161-G2/G3-2	S06G2/G3-2	ACGGTGAGGAGTGGCACATG[A/C]GATGGAAAGTTCCTGTAGAC
UMC161	UMC161-G2/G3-2B	S06G2/G3-2B	ACGGTAAGGAGTGGCACATG[A/C]GATGGAAAGTTCCTGTAGAC
UMC107	UMC107G2/G4-1	S05G2/G4-1	TATGCTTGGAAAGTGGGAAA[G/]GGGAACATACGATGGAGGAC
UMC67	UMC67-G5/G6-1	S03G5/G6-1	AAACAATAATTTTTACACAG[/T]TGCTAAGGTTTTACTGTTTT
UMC67	UMC67-G2/G6-1	S03G2/G6-1	ATATCCATGTTGTCGCCTGC[/TG]TGTGCGCTTGCTTGCCGCTA
UMC76	UMC76-G4/G6-1	S02G4/G6-1	TTGCTGCTATGTTTACTGGG[/T]TGTAGAAAAAAAAATAATAT
UMC76	UMC76-G2/G6-1	S02G2/G6-1	GCTCGGTAATAATTCTGGCT[C/G]CGATGGCACCCATATTCCTC
UMC76	UMC76-G2/G6-1B	S02G2/G6-1B	GCTCGGTAATAATTCTGGCT[C/G]CGATGGCACCCATATTCCTG
UMC76	UMC76-G2/G5-1	S02G2/G5-1	AAAACACGTGGTGTTTGTTA[G/A]GAAAGACCTAGTTTCTCGGC
UMC76	UMC76-G2/G5-1B	S02G2/G5-1B	AAATCACGTGGTGTTTGTTA[G/A]GAAAGACCTAGTTTCTCGGC
UMC76	UMC76-G2/G5-1	S02G2/G5-1	TAGTTTCTCGGCAATTGGCA[G/T]TGTGGAATGACCATCTCGTG
UMC76	UMC76-G2/G5-1B	S02G2/G5-1B	TAGTTTCTCGGCAATTGGCA[G/T]TGTGGAATGACCATCTCGTC
UMC76	UMC76-G2/G5-2	S02G2/G5-2	GTGTGGAATGACCATCTCGT[G/C]GTGATGCCAGCATGCTGTTA
UMC76	UMC76-G2/G5-2B	S02G2/G5-2B	GTGTGGAATGACCATCTCGT[G/C]GTGATGCCAGCATGCTACTA
UMC76	UMC76-G2/G5-3	S02G2/G5-3	ACCCTGTCAGGCTTCCACAG[A/C]TATAATATTTGTTGTGGTGT
UMC76	UMC76-G2/G5-3B	S02G2/G5-3B	ACTCTGTCAGGCTTCCACAG[A/C]TATAATATTTGTTGTGGTGT
UMC76	UMC76-G2/G5-3C	S02G2/G5-3C	ACTCTGTCAGGCTTCCACAG[A/C]TATAATATTTGTTGTGTGTG
UMC76	UMC76-G2/G5-3D	S02G2/G5-3D	ACCCTGTCAGGCTTCCACAG[A/C]TATAATATTTGTTGTGTGTG

Example 4 Analysis of Polymorphisms

A. Preparation of Samples

Polymorphisms are detected in a target nucleic acid from a plant being analyzed. Target nucleic acids can be genomic or cDNA. Many of the methods described below require amplification of DNA from target samples. This can be accomplished by e.g., PCR. *See generally PCR Technology: Principles and Applications for DNA Amplification* (ed. H.A. Erlich, Freeman Press, NY, NY, 1992); *PCR Protocols: A Guide to Methods and Applications* (eds. Innis, et al., Academic Press, San Diego, CA, 1990); Mattila et al., *Nucleic Acids* Res. 19, 4967 (1991); Eckert et al., *PCR Methods and Applications* 1, 17 (1991); *PCR* (eds. McPherson et al., IRL Press, Oxford); and U.S. Patent 4,683,202 (each of which is incorporated by reference for all purposes).

Other suitable amplification methods include the ligase chain reaction (LCR) (see Wu and Wallace, *Genomics* 4, 560 (1989), Landegren et al., Science 241, 1077 (1988), transcription amplification (Kwoh et al., *Proc. Natl. Acad.* Sci. USA 86, 1173 (1989)), and self-sustained sequence replication (Guatelli et al., *Proc. Nat. Acad. Sci.* USA, 87, 1874 (1990)) and nucleic acid based sequence amplification (NASBA). The latter two amplification methods involve isothermal reactions based on isothermal transcription, which produce both single stranded RNA (ssRNA) and double stranded DNA (dSDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

B. Detection of Polymorphisms in Target DNA

There are two distinct types of analysis depending whether a polymorphism in question has already been characterized. The first type of analysis is sometimes referred to as de novo characterization. This analysis compares target sequences in different individual plants to identify points of variation, i.e., polymorphic sites. The de novo identification of the polymorphisms of the invention is described in the Examples section, The second type of

analysis is determining which form(s) of a characterized polymorphism is (are) present in plants under test. There are a variety of suitable procedures, which are discussed in turn.

1. Allele-Specific Probes

The design and use of allele-specific probes for analyzing polymorphisms is described by e.g., Saiki et al., *Nature* 324, 163-166 (1986); Dattagupta, EP 235,726, Saiki, WO 89/11548. Allele-specific probes can be designed that hybridize to a segment of target DNA from one member of a species but do not hybridize to the corresponding segment from another member due to the presence of different polymorphic forms in the respective segments from the two members. Hybridization conditions should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles. Some probes are designed to hybridize to a segment of target DNA such that the polymorphic site aligns with a central position (e.g., in a 15 mer at the 7 position; in a 16 mer, at either the 8 or 9 position) of the probe. This design of probe achieves good discrimination in hybridization between different allelic forms.

Allele-specific probes are often used in pairs, one member of a pair showing a perfect match to a reference form of a target sequence and the other member showing a perfect match to a variant form. Several pairs of probes can then be immobilized on the same support for simultaneous analysis of multiple polymorphisms within the same target sequence.

2. Tiling Arrays

The polymorphisms can also be identified by hybridization to nucleic acid arrays, some example of which are described by Wo 95/11995 (incorporated by reference in

its entirety for all purposes). One form of such arrays is described in the Examples section in connection with de novo identification of polymorphisms. The same array or a different array can be used for analysis of characterized polymorphisms. WO 95/11995 also describes subarrays that are optimized for detection of a variant forms of a precharacterized polymorphism. Such a subarray contains probes designed to be complementary to a second reference sequence, which is an allelic variant of the first reference sequence. The second group of probes is designed by the same principles as described in the Examples except that the probe" exhibit complementarity to the second reference sequence. The inclusion of a second group (or further groups) can be particular useful for analysing short subsequences of the primary reference sequence in which multiple mutations are expected to occur within a short distance commensurate with the length of the probes (*i.e.*, two or more mutations within 9 to 21 bases).

3. Allele-Specific Primers

An allele-specific primer hybridizes to a site on target DNA overlapping a polymorphism and only primes amplification of an allelic form to which the primer exhibits perfect complementarity. See Gibbs, *Nucleic Acid Res . 17* , 2427-2448 (1989). This primer is used in conjunction with a second primer which hybridizes at a distal site. Amplification proceeds from the two primers leading to a detectable product signifying the particular allelic form is present. A control is usually performed with a second pair of primers, one of which shows a single base mismatch at the polymorphic site and the other of which exhibits perfect complementarity to a distal site. The single-base mismatch prevents amplification and no detectable product is formed. The method works best when the mismatch is included in the 3'-most position of the oligonucleotide aligned with the polymorphism because this position is most destabilizing to elongation from the primer. *See, e.g.*, WO 93/22456.

4. Direct-Sequencing

The direct analysis of the sequence of polymorphisms of the present invention can bo accomplished using either the dideoxy chain termination method or the Maxam Gilbert method (see Sambrook et al., *Molecular Cloning, A Laboratory Manual* (2nd Ed., CSHP, New York 1989); Zyskind et al., *Recombinant DNA Laboratory Manual*, (Acad. Press, 1988)).

5. Denaturing Gradient Gel Electrophoresis

Amplification products generated using the polymerase chain reaction can be analyzed by the use of denaturing gradient gel electrophoresis. Different alleles can be identified based on the different sequence-dependent melting properties and electrophoretic migration of DNA in solution, Erlich, ed., *PCR Technology, Principles and Applications for DNA Amplification*, (W. H. Freeman and Co, New York, 1992), Chapter 7.

6. Single-Strand Conformation Polymorphism Analysis

Alleles of target sequences can be differantiated using single-strand conformation polymorphism analysis, which identifies base differences by alteration in electrophoretic migration of single stranded PCR products, as described in Orita et al., *Proc, Nat. Acad. Sci.* 86, 2766-2770 (1989). Amplified PCR products can be generated as described above, and heated or otherwise denatured, to form single stranded amplification products. Single-stranded nucleic acids may refold or form secondary structures which are partially dependent on the base sequence. The different electrophoretic mobilities of single-stranded amplification products can be related to base-sequence difference between alleles of target sequences.

Example 5 . Methods of Use

After determining polymorphic form(s) present in a subject plant at one or more polymorphic sites, this information can be used in a number of methods.

A. Fingerprint Analysis

Analysis of which polymorphisms are present in a plant is useful in determining of which strain the plant is a member and in distinguishing one strain from another. A genetic fingerprint for an individual strain can be made by determining the nucleic acid sequence possessed by that individual strain that corresponds to a region of the genome known to contain polymorphisms. For a discussion of genetic fingerprinting in the animal kingdom, see, for example, Stokening et.al., *Am. J. Hum. Genet.* 48:370-382 (1991). The probability that one or more polymorphisms in an individual strain is the same as that in any other individual strain decreases as the number of polymorphic sites is increased.

The comparison of the nucleic acid sequences from two strains at one or multiple polymorphic sites can also demonstrate common or disparate ancestry. Since the polymorphic sites are within a large region in the genome, the probability of recombination between these polymorphic sites is low. That low probability means the haplotype (the set of all the disclosed polymorphic sites) set forth in this application should be inherited without change for at least several generations. Knowledge of plant strain or ancestry is useful, for example, in a plant breeding program or in tracing progeny of a proprietary plant. Fingerprints are also used to identify an individual strain and to distinguish or determine the relatedness of one individual strain to another. Genetic fingerprinting can also be useful in hybrid certification, the certification of seed lots, and the assertion of plant breeders rights under the laws of various countries.

B. Correlation of Polymorphisms with Phenotypic Traits

The polymorphisms of the invention may contribute to the phenotype of a plant in different ways. Some polymorphisms occur within a protein coding sequence and contribute to phenotype by affecting protein structure. The effect may be neutral, beneficial or detrimental, or both beneficial and detrimental, depending on the circumstances. Other polymorphisms occur in noncoding regions but may exert phenotypic effects indirectly via influence on replication, transcription, and translation. A single polymorphism may affect more than one phenotypic trait. Likewise, a single phenotypic trait may be affected by polymorphisms in different genes. Further, some polymorphisms predispose a plant to a distinct mutation that is causally related to a certain phenotype.

Phenotypic traits include characteristics such as growth rate, crop yield, crop quality, resistance to pathogens, herbicides, and other toxins, nutrient requirements, resistance to high temperature, freezing, drought, requirements for light and soil type, aesthetics, and height. Other phenotypic traits include susceptibility or resistance to diseases, such as plant cancers. Often polymorphisms occurring within the same gene correlate with the same phenotype.

Correlation is performed for a population of plants, which have been tested for the presence or absence of a phenotypic trait of interest and for polymorphic markers sets. To perform such analysis, the presence or absence of a set of polymorphisms (i.e. a polymorphic set) is determined for a set of the plants, some of whom exhibit a particular trait, and some of which exhibit lack of the trait. The alleles of each polymorphism of the set are then reviewed to determine whether the presence or absence of a particular allele is associated with the trait of interest. Correlation can be performed by standard statistical methods such as a K-squared test and statistically significant correlations between polymorphic form(s) and phenotypic characteristics are noted.

Correlations between characteristics and phenotype are useful for breeding for desired characteristics. By analogy, Beitz et al., US 5,292,639 discuss use of bovine mitochondrial polymorphisms in a breeding program to improve milk production in cows. To evaluate the effect of mtDNA D-loop sequence polymorphism on milk production, each cow was assigned a value of 1 if variant or 0 if wildtype with respect to a prototypical mitochondrial DNA sequence at each of 17 locations considered. Each production trait was analyzed individually with the following animal model:

$$Y_{ijkpn} = \mu + YS_i + Pj; + X_k \sim \beta_1 + \ldots \beta_{17} + PE_n + a_n + e_p$$

where Y_{ijkpn} is the milk, fat, fat percentage, SNF, SNF percentage, energy concentration, or lactation energy record; μ is an overall mean; YS_i is the effect common to all cows calving in year-season; X_k is the effect common to cows in either the high or average selection line; β_1 to β_{17} are the binomial regressions of production record on mtDNA D-loop sequence polymorphisms; PE_n is permanent environmental effect common to all records of cow n; a_n is effect of animal n and is composed of the additive genetic contribution of sire and dam breeding values and a Mendelian sampling effect; and e_p is a random residual. It was found that eleven of seventeen polymorphisms tested influenced at least one production trait. Bovines having the best polymorphic forms for milk production at these eleven loci are used as parents for breeding the next generation of the herd.

One can test at least several hundreds of markers simultaneously in order to identify those linked to a gene or chromosomal region. For example, to identify markers linked to a gene conferring disease resistance, a DNA pool is constructed from plants of a segregating population that are resistant and another pool is constructed from plants that are sensitive to the disease. Those two DNA pools are identical except for the DNA sequences at the resistance gene locus and in the surrounding genomic area. Hybridization of such DNA pools to the DNA sequences listed in Table 1 allows

the simultaneous testing of several hundreds of loci for polymorphisms. Allelic polymorphism-detecting sequences that show differences in hybridization patterns between such DNA pools will represent loci linked to the disease resistance gene.

The method just described can also be applied to rapidly identify rare alleles in large populations of plants. For example, nucleic acid pools are constructed from several individuals of a large population. The nucleic acid pools are hybridized to nucleic acids having the polymorphism-detecting sequences listed in Table I. The detection of a rare hybridization profile will indicate the presence of a rare allele in a specific nucleic acid pool. RNA pools are particularly suited to identify differences in gene expression.

C. Marker assisted back-cross

The markers are used to select, in back-cross populations, the plant that have the higher percentage of recurrent parent, while still remaining the genes given by the donor plant.

Example 6. Modified Polypeptides and Gene Sequences

The invention further provides variant forms of nucleic acids and corresponding proteins. The nucleic acids comprise at least 10 contiguous amino acids of one of the sequences for example as described in Table I, in any of the allelic forms shown. Some nucleic acid encode full-length proteins.

Genes can be expressed in an expression vector in which a gene is operably linked to a native or other promoter. Usually, the promoter is an eukaryotic promoter for expression in a eukaryotic cell. The transcription regulation sequences typically include an heterologous promoter and

optionally an enhancer which is recognized by the host. The selection of an appropriate promoter, for example trp, lac, phage promoters, glycolytic enzyme promoters and tRNA promoters, depends on the host selected. Commercially available expression vectors can be used. Vectors can include host-recognized replication systems, amplifiable genes, selectable markers, host sequences useful for insertion into the host genome, and the like.

The means of introducing the expression construct into a host cell varies depending upon the particular construction and the target host. Suitable means include fusion, conjugation, transfection, transduction, electroporation or injection, as described in Sambrook, *supra*. A wide variety of host cells can be employed for expression of the variant gene, both prokaryotic and eukaryotic. Suitable host cells include bacteria such as *E. coli*, yeast, filamentous fungi, insect cells, mammalian cells, typically immortalized, e.g., mouse, CHO, human and monkey cell lines and derivatives thereof, and plant cells. Preferred host cells are able to process the variant gene product to produce an appropriate mature polypeptide. Processing includes glycosylation, ubiquitination, disulfide bond formation, general post-translational modification, and the like.

The DNA fragments are introduced into cultured plant cells by standard methods including electroporation (From et al., *Proc. Natl Acad. Sci*, USA 82, 5824 (19853, infection by viral vectors such as cauliflower mosaic virus (CaMV) (Hohn et *al.* , *Molecular Biology of Plant Tumors*, (Academic Press, New York, 1982) pp. 549-560; Howell, US 4,407,956), high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface (Klein et al., *Nature 327*, 70-73 (1987)), USQ of pollen as vector (WO 85/01856), or use of *Agrobacterium tumefaciens* transformed with a Ti plasmid in which DNA fragments are cloned. The Ti plasmid is transmitted to plant cells upon infection by

Agrobacterium tumefaciens, and is stably integrated into the plant genome (Horsch et al., *Science*, 233, 496-498 (1984); Fraley et al., *Proc. Natl. Acad. Sci. USA* 80, 4803 (1983)).

The protein may be isolated by conventional means of protein biochemistry and purification to obtain a substantially pure product, *i.e.*, 80, 95 or 99% free of cell component contaminants, as described in Jacoby, *Methods in Enzymology* Volume 104, Academic Press, New York (1984); Sc:opes, *Protein Purification, Principles and Practice'*, 2nd Edition, Springer-Verlag, New York (1987); and Deutscher (ed), *Guide to Protein Purifl cation' Methods in Enzymology*, Vol. 182 (1990). If the protein is secreted, it can be isolated from the supernatant in which the host cell is grown. If not secreted, the protein can be isolated from a lysate of the host cells.

The invention further provides transgenic plants capable of expressing an exogenous variant gene and/or having one or both alleles of an endogenous variant gene inactivated. Plant regeneration from cultural protoplasts is described in Evans et al., "Protoplasts Isolation and Culture," *Handbook of Plant Cell Cultures 1* , 124-176 (MacMillan Publishing Co., New York, 1983); Davey, "Recent Developments in the Culture and Regeneration of Plant Protoplasts," *Protoplasts, (1983)* - pp. 12-29, (Birkhauser, Basal 1983); Dale, "Protoplast Culture and Plant Regeneration of Cereals and Other Recalcitrant Crops," *Protoplasts (1983)* - pp. 31-41, (Birkhauser, Basel 1983); Binding, "Regeneration of Plants," *Plant ProtopLasts, pp. 21*-73, (CRC Press, Boca Raton, 1985). For example, a variant gene responsible for a disease-resistant phenotype can be introduced into the plant to simulate that phenotype. Expression of an exogenous variant gene is usually achieved by operably linking the qene to a promoter and optionally an enhancer. Inactivation of an exogenous variant genes can be achieved by forming a transgene in which a cloned variant genes is inactivated by insertion of a positive selection marker. *See* Capecchi, *Science* 244, 1288-1292 (1989). Such transgenic plant are

useful in a variety of screening assays. For example, the transgenic plant can then be treated with compounds of interest and the effect of those compounds on the disease resistance can be monitored. In another example, the transgenic plant can be exposed to a variety of environmental conditions to determine the effect of those conditions on the resistance to the disease.

In addition to substantially full-length polypeptides, the present invention includes biologically active fragments of the polypeptides, or analogs thereof, including organic molecules which simulate the interactions of the peptides. Biologically active fragments include any portion of the full-length polypeptide which confers a biological function on the variant gene product, including ligand binding, and antibody binding. Ligand binding includes binding by nucleic acids, proteins or polypeptides, small biologically active molecules, or large cellular structures.

Polyclonal and/or monoclonal antibodies that specifically bind to one allelic gene products but not to a second allelic gene product are also provided. Antibodies can be made by injecting mice or other animals with the variant gene product or synthetic peptide fragments thereof. Monoclonal antibodies are screened as are described, for example, in Harlow & Lane, *Antibodies, A Laboratory Manual*, Cold Spring Harbor Press, New York (1988); Goding, *Monoclonal antibodies, Principles* and Practice (2d ed.) Academic Press, New York (1986). Monoclonal antibodies are tested for specific immunoreactivity with a variant gene product and lack of immunoreactivity to the corresponding prototypical gene product. These antibodies are useful in diagnostic assays for detection of the variant form, or as an active ingredient in a pharmaceutical composition.

Example 7. Kits

The invention further provides kits comprising at least one allele-specific oligonucleotide as described above. Often, the kits contain one or more pairs of

allele-specific oligonucleotides hybridizing to different forms of a polymorphism. In some kits, the allele-specific oligonucleotides are provided immobilized to a substrate. For example, the same substrate can comprise allele-specific oligonucleotide probes for detecting at least 10, 100 or all of the polymorphisms shown in Table I. Optional additional components of the kit include, for example, restriction enzymes, reverse-transcriptase or polymerase, the substrate nucleoside triphosphates, means used to label (for example, an avidin-enzyme conjugate and enzyme substrate and chromogen if the label is biotin), and the appropriate-buffers for reverse transcription, PCR, or hybridization reactions. Usually, the kit also contains instructions for carrying out the methods.

CLAIMS

1. A nucleic acid segment comprising at least 10 contiguous nucleotides from a vegetal sequence including a polymorphic site, notably a Single Nucleotide Polymorphism (SNP) or the complement of the segment.

2. A nucleic acid segment of claim 1, which is comprised in the sequence shown in Table I.

3. A nucleic acid segment of claim 1, less than 100 bases.

4. A nucleic acid segment of claim 1, that is DNA.

5. A nucleic acid segment of claim 1, that is RNA

6. The segment of claim 1 that is less than 50 bases.

7. The segment of claim 1, that is less than 20 bases.

8. An allele-specific oligonucleotide that hybridizes to a sequence of claim 1 or its complement.

9. An allele-specific oligonucleotide that hybridizes to a sequence of claim 8, sequence shown in Table 1.

10. The allele-specific oligonucleotide of claim 8, that is a probe

11. The allele-specific oligonucleotide of claim 10, wherein the central position of the probe aligns with the polymorphic site in the sequence.

12. The allele-specific oligonucleotide of claim 8, that is a primer.

13. The allele-specific oligonucleotide of claim 12, primer which comprises a sequence shown in Table I

14. The allele-specific oligonucleotide of claim 12, 3' end primer which comprises a sequence shown in Table I.

15. The method of analysing a nucleic acid, comprising : obtaining the nucleic acid from a subject; and

determining a base occupying any one of the polymorphic sites shown in Table I.

16. The method of claim 15, wherein the determining comprises determining a set of bases occupying a set of the polymorphic sites shown in Table I.

17. The method of claim 16, wherein the nucleic acid is obtained from a plurality of subjects, and a base occupying one of the polymorphic positions is determined in each of the subjects, and the method further comprises testing each subject for the presence of a phenotype, and correlating the presence of the phenotype with the base.

18. Kit comprising at least one allele-specific oligonucleotide of claim 1 and optional additional composants (enzymes, buffers, instructions...)

19. Kit according to claim 18 comprising at least one allele-specific oligonucleotide of claim 2.

20 Use of the nucleic segments according to claims 1 to 17, to demonstrate common or disparate ancestry.

21. Use of the nucleic segments according to claims 1 to 17 in plant breeding.

22. Use of the nucleic acid segments according to claims 1 to 17 to trace progeny of a priority plant.

23. Use of the nucleic acid segments according to claims 1 to 17 in hybrid certification.

24. Use of the nucleic acid segments according to claims 1 to 17 to select in a back-cross population the plants that have the higher percentage of recurrent parent (marker assisted back-cross).

25. Use of the nucleic segments according to claim 1 to 17, wherein the polymorphisms, all of them or most of them, are linked to a group of genes involved in a given metabolic pathway.

26. Use according to 25, wherein the metabolic pathway is selected from the oil metabolic pathway, the starch metabolic pathway, the protein metabolic pathway, the aminoacids metabolic pathway, the lignin and the cell wall

composition metabolic pathway and the pathogene resistance pathway

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68

A3

(11) International Publication Number: **WO 98/30717**

(43) International Publication Date: 16 July 1998 (16.07.98)

(21) International Application Number: PCT/EP97/07134

(22) International Filing Date: 2 December 1997 (02.12.97)

(30) Priority Data:
60/032,069 2 December 1996 (02.12.96) US

(71) Applicant *(for all designated States except US)*: BIOCEM S.A. [FR/FR]; Campus Universitaire des Cézeaux, 24, avenue des Landais, F–63170 Aubière (FR).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: MURIGNEUX, Alain [FR/FR]; Biocem S.A., Campus Universitaire des Cézeaux, 24, avenue des Landais, F–63170 Aubière (FR).

(74) Agent: BREESE–MAJEROWICZ; 3, avenue de l'Opéra, F–75001 Paris (FR).

(81) Designated States: AU, CA, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report:
29 April 1999 (29.04.99)

(54) Title: VEGETAL SEQUENCES INCLUDING A POLYMORPHIC SITE AND THEIR USES

(57) Abstract

A nucleic acid segment comprising at least 10 contiguous nucleotides from a vegetal sequence including a polymorphic site; or the complement of the segment.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 97/07134

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SHATTUCK-EIDENS D M ET AL.: "Rapid detection of maize DNA sequence varition" GENETIC ANALYSIS - TECHNIQUES AND APPLICATIONS, vol. 8, no. 8, 1991, pages 240-245, XP002085038 see abstract	1,4,8, 10,12,20
Y	see page 240, column 1, paragraph 1 - page 242, column 1, paragraph 1; figures 1,3 --- -/--	2,9, 15-19

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 November 1998	16. 03. 1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Knehr, M

2

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BECKMANN J S: "Oligonucleotide polymorphisms:A new tool for genomic genetics" BIO/TECHNOLOGY, vol. 6, 1988, pages 1061-1064, XP002085039	1,4,8, 20,21,24
Y	see the whole document	2,9, 15-17

X	PLASCHKE J ET AL: "DETECTION OF GENETIC DIVERSITY IN CLOSELY RELATED BREAD WHEAT USING MICROSATELLITE MARKERS" THEORETICAL AND APPLIED GENETICS, vol. 91, no. 6/07, November 1995, pages 1001-1007, XP000604133 see abstract see page 1001, column 1, paragraph 1 - page 1003, column 1, paragraph 4 see page 1006, column 1, paragraph 4 - page 1007, column 1, paragraph 1; figure 2; table 2	1,4,8, 12,20,21

X	POWELL W ET AL: "HYPERVARIABLE MICROSATELLITES PROVIDE A GENERAL SOURCE OF POLYMORPHIC DNA MARKERS FOR THE CHLOROPLAST GENOME" CURRENT BIOLOGY, vol. 5, no. 9, 1 September 1995, pages 1023-1029, XP000570201 see abstract see page 1023, column 1, paragraph 1 - page 1026, column 2, paragraph 1; figure 2; tables 1,3	1,4,20

X	SENIOR M L ET AL: "MAPPING MAIZE MICROSATELLITES AND POLYMERASE CHAIN REACTION CONFIRMATION OF THE TARGETED REPEATS USING A CT PRIMER" GENOME, vol. 36, no. 5, 1 October 1993, pages 884-889, XP000569589 see abstract see page 885, column 1, paragraph 2 - column 2, paragraph 3; figure 1; table 1	1,4,8

	-/--	

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HANSON M A ET AL.: "Evolution of anthocyanin biosynthesis in maize kernels: The role of regulatory and enzymatic loci" GENETICS, vol. 143, 1996, pages 1395-1407, XP002085040 see abstract see page 1395, column 1, paragraph 1 - column 2, paragraph 2 see page 1396, column 2, paragraph 2 - column 2, paragraph 2 see page 1398, column 2, paragraph 2 - paragraph 3; figure 2 ---	1,4,25, 26
X	SHATTUCK-EIDENS D M ET AL.: "DNA sequence variation within maize and melon: Observations from polymerase chain reaction amplification and direct sequencing" GENETICS, vol. 126, 1990, pages 207-217, XP002085041 see abstract see page 207, column 1, paragraph 1 - page 208, column 2, paragraph 4 see page 211, column 1, paragraph 2 - column 2, paragraph 1; figures 1,2; tables 3-7 ---	1,4,8,12
X	US 5 437 697 A (SEBASTIAN SCOTT A ET AL) 1 August 1995 * see especially column 17, lines 4 to 33 * see the whole document ---	1,4, 20-24
X	US 5 332 408 A (METS LAURENS J ET AL) 26 July 1994 * see especially example 1 * see the whole document ---	1,4, 20-24
X	WO 92 07948 A (LUBRIZOL CORP) 14 May 1992	1,3,4, 6-8,12
Y	see the whole document ---	18,19
Y	WO 89 07647 A (PIONEER HI BRED INT) 24 August 1989 see abstract see page 1, paragraph 1 - page 4, paragraph 1 see page 11, paragraph 4; claims 1,2,6,7; table 1 ---	1,4,8, 12,21-23
Y	EP 0 317 239 A (NATIVE PLANTS INC) 24 May 1989 see the whole document ---	1,4,8, 12,21-23
	-/--	

2

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 97/07134

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	EP 0 785 281 A (SAPPORO BREWERIES) 23 July 1997 see the whole document ---	1,3,4, 6-8,12, 18,20
E	WO 98 24796 A (LANDRY BENOIT S ;LEMIEUX BERTRAND (CA); MURIGNEUX ALAIN (FR); SAPO) 11 June 1998 see the whole document ---	1,3-8, 10-14, 18,20-24
E	WO 98 30721 A (PIONEER HI BRED INT ;BIRO RONALD L (US); FEAZEL RHONDA (US); HELEN) 16 July 1998 see the whole document ---	1,3-8, 10,12
T	WO 98 30717 A (BIOCEM S A ;MURIGNEUX ALAIN (FR)) 16 July 1998 see the whole document -----	1-26

2

INTERNATIONAL SEARCH REPORT

Inte tional application No.
PCT/EP 97/07134

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see FURTHER INFORMATION sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

see FURTHER INFORMATION sheet, subject 1.

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-26 (partial)

INVENTION 1:
A nucleic acid segment from a vegetal sequence including a polymorphic site, and in particular SEQ ID NOs: 67 and 68 (Bt2 gene/marker from maize), an allele-specific oligonucleotide hybridizing to such sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

2. Claims: 1-26 (partial)

INVENTION 2:
SEQ ID NOs: 69 to 76 (Ssu gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

3. Claims: 1-26 (partial)

INVENTION 3:
SEQ ID NOs: 77 to 82 (Bt1 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

4. Claims: 1-26 (partial)

INVENTION 4:
SEQ ID NOs: 83 to 90 (Bre1 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

5. Claims: 1-26 (partial)

INVENTION 5:
SEQ ID NOs: 91 to 104 (ASG12 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

6. Claims: 1-26 (partial)

INVENTION 6:
SEQ ID NOs: 105 to 114 (Sh2 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

7. Claims: 1-26 (partial)

INVENTION 7:
SEQ ID NOs: 115 to 132 (Sh1 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

8. Claims: 1-26 (partial)

INVENTION 8:
SEQ ID NOs: 133 to 144 (UAZ77 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

9. Claims: 1-26 (partial)

INVENTION 9:
SEQ ID NOs: 145 and 146 (UAZ171 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

10. Claims: 1-26 (partial)

INVENTION 10:
SEQ ID NOs: 147 to 150 (UMC17 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

11. Claims: 1-26 (partial)

INVENTION 11:
SEQ ID NOs: 151 to 178 (CSU109 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these

sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

12. Claims: 1-26 (partial)

INVENTION 12:
SEQ ID NOs: 179 to 180 (UMC130 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

13. Claims: 1-26 (partial)

INVENTION 13:
SEQ ID NOs: 181 to 212 (CSU61 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

14. Claims: 1-26 (partial)

INVENTION 14:
SEQ ID NOs: 213 to 234 (UMC95 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

15. Claims: 1-26 (partial)

INVENTION 15:
SEQ ID NOs: 235 to 290 (Wx1 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

16. Claims: 1-26 (partial)

INVENTION 16:
SEQ ID NOs: 291 to 300 (UMC109 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

17. Claims: 1-26 (partial)

INVENTION 17:
SEQ ID NOs: 301 to 320 (UMC80 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

18. Claims: 1-26 (partial)

INVENTION 18:
SEQ ID NOs: 321 to 358 (UMC254 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

19. Claims: 1-26 (partial)

INVENTION 19:
SEQ ID NOs: 359 to 366 (ASG49 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

20. Claims: 1-26 (partial)

INVENTION 20:
SEQ ID NOs: 367 to 370 (ASG8 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

21. Claims: 1-26 (partial)

INVENTION 21:
SEQ ID NOs: 371 to 374 (UMC132 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

22. Claims: 1-26 (partial)

INVENTION 22:
SEQ ID NOs: 375 to 406 (UMC21 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these

sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

23. Claims: 1-26 (partial)

INVENTION 23:
SEQ ID NOs: 407 to 416 (UMC65 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

24. Claims: 1-26 (partial)

INVENTION 24:
SEQ ID NOs: 417 to 434 (UMC59 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

25. Claims: 1-26 (partial)

INVENTION 25:
SEQ ID NOs: 435 to 450 (Ac1 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

26. Claims: 1-26 (partial)

INVENTION 26:
SEQ ID NOs: 451 to 456 (UMC90 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

27. Claims: 1-26 (partial)

INVENTION 27:
SEQ ID NOs: 457 to 460 (UMC66 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

28. Claims: 1-26 (partial)

INVENTION 28:
SEQ ID NOs: 461 to 464 (Adh2 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

29. Claims: 1-26 (partial)

INVENTION 29:
SEQ ID NOs: 465 to 482 (UMC63 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

30. Claims: 1-26 (partial)

INVENTION 30:
SEQ ID NOs: 483 to 486 (UMC102 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

31. Claims: 1-26 (partial)

INVENTION 31:
SEQ ID NOs: 487 to 490 (ASG24 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

32. Claims: 1-26 (partial)

INVENTION 32:
SEQ ID NOs: 491 to 522 (UMC49 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

33. Claims: 1-26 (partial)

INVENTION 33:
SEQ ID NOs: 523 to 534 (UMC131 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these

sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

34. Claims: 1-26 (partial)

INVENTION 34:
SEQ ID NOs: 535 to 552 (UMC53 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

35. Claims: 1-26 (partial)

INVENTION 35:
SEQ ID NOs: 553 to 558 (UMC161 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

36. Claims: 1-26 (partial)

INVENTION 36:
SEQ ID NOs: 559 and 560 (UMC107 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

37. Claims: 1-26 (partial)

INVENTION 37:
SEQ ID NOs: 561 to 564 (UMC67 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

38. Claims: 1-26 (partial)

INVENTION 38:
SEQ ID NOs: 565 to 590 (UMC76 gene/marker from maize), an allele-specific oligonucleotide hybridizing to these sequences or their complements, a method of analyzing them, a kit containing such allele-specific oligonucleotides, and the use of such sequences.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/EP 97/07134

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5437697 A	01-08-1995	US 5746023 A	05-05-1998
US 5332408 A	26-07-1994	NONE	
WO 9207948 A	14-05-1992	AU 8953991 A	26-05-1992
		CA 2073184 A	07-05-1992
		EP 0509089 A	21-10-1992
WO 8907647 A	24-08-1989	AU 631562 B	03-12-1992
		AU 4030289 A	06-09-1989
		EP 0402401 A	19-12-1990
EP 0317239 A	24-05-1989	CA 1323553 A	26-10-1993
		JP 2002400 A	08-01-1990
		JP 2634208 B	23-07-1997
		US 5324631 A	28-06-1994
EP 0785281 A	23-07-1997	AU 6531796 A	26-02-1997
		CZ 9700986 A	17-09-1997
		SK 41797 A	08-10-1997
		CA 2201127 A	13-02-1997
		CN 1159212 A	10-09-1997
		WO 9705281 A	13-02-1997
WO 9824796 A	11-06-1998	AU 5511598 A	29-06-1998
		AU 7206698 A	03-08-1998
		WO 9830717 A	16-07-1998
WO 9830721 A	16-07-1998	AU 6024598 A	03-08-1998
WO 9830717 A	16-07-1998	AU 5511598 A	29-06-1998
		AU 7206698 A	03-08-1998
		WO 9824796 A	11-06-1998